HHS Public Access

Author manuscript

N Engl J Med. Author manuscript; available in PMC 2019 September 07.

Published in final edited form as:

N Engl J Med. 2019 March 07; 380(10): 905–914. doi:10.1056/NEJMoa1810642.

Patient Safety Outcomes under Flexible and Standard Resident Duty-Hour Rules

Jeffrey H. Silber, M.D., Ph.D., **Lisa M. Bellini, M.D.**, **Judy A. Shea, Ph.D.**, **Sanjay V. Desai, M.D.**, **David F. Dinges, Ph.D.**, **Mathias Basner, M.D., Ph.D.**, **Orit Even-Shoshan, M.S.**, **Alexander S. Hill, B.S.**, **Lauren L. Hochman, B.A.**, **Joel T. Katz, M.D.**, **Richard N. Ross, M.S.**, **David M. Shade, J.D.**, **Dylan S. Small, Ph.D.**, **Alice L. Sternberg, Sc.M.**, **James Tonascia, Ph.D.**, **Kevin G. Volpp, M.D., Ph.D.**, **David A. Asch, M.D., M.B.A.**, and **iCOMPARE Research Group***

Center for Outcomes Research, Children's Hospital of Philadelphia (J.H.S., O.E.-S., A.S.H., L.L.H., R.N.R.), the Departments of Pediatrics (J.H.S.), Anesthesiology and Critical Care (J.H.S.), and Medicine (L.M.B., J.A.S., K.G.V., D.A.A.), University of Pennsylvania School of Medicine, the Departments of Health Care Management (J.H.S., K.G.V., D.A.A.) and Statistics (D.S.S.), the Wharton School, the Leonard Davis Institute of Health Economics (J.H.S., J.A.S., D.S.S., K.G.V., D.A.A.), and the Department of Psychiatry (D.F.D., M.B.), University of Pennsylvania, and the Corporal Michael J. Crescenz Veterans Affairs Medical Center (K.G.V., D.A.A.) — all in Philadelphia; the Departments of Medicine (S.V.D.), Epidemiology (D.M.S., A.L.S., J.T.), and Biostatistics (J.T.), Johns Hopkins University, Baltimore; and the Department of Medicine, Brigham and Women's Hospital, Boston (J.T.K.). Address reprint requests to Dr. Silber at the Center for Outcomes Research, Children's Hospital of Philadelphia, Roberts Center, 2716 South St., Rm. 5123, Philadelphia, PA 19146-2305, or at silber@email.chop.edu.

Abstract

BACKGROUND—Concern persists that extended shifts in medical residency programs may adversely affect patient safety.

METHODS—We conducted a cluster-randomized noninferiority trial in 63 internal-medicine residency programs during the 2015–2016 academic year. Programs underwent randomization to a group with standard duty hours, as adopted by the Accreditation Council for Graduate Medical Education (ACGME) in July 2011, or to a group with more flexible duty-hour rules that did not specify limits on shift length or mandatory time off between shifts. The primary outcome for each program was the change in unadjusted 30-day mortality from the pretrial year to the trial year, as ascertained from Medicare claims. We hypothesized that the change in 30-day mortality in the flexible programs would not be worse than the change in the standard programs (difference-in-difference analysis) by more than 1 percentage point (noninferiority margin). Secondary outcomes were changes in five other patient safety measures and risk-adjusted outcomes for all measures.

*A complete list of the members of the iCOMPARE Research Group is provided in the Supplementary Appendix, available at NEJM.org.

A data sharing statement provided by the authors is available with the full text of this article at NEJM.org.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

RESULTS—The change in 30-day mortality (primary outcome) among the patients in the flexible programs (12.5% in the trial year vs. 12.6% in the pretrial year) was noninferior to that in the standard programs (12.2% in the trial year vs. 12.7% in the pretrial year). The test for noninferiority was significant (P = 0.03), with an estimate of the upper limit of the one-sided 95% confidence interval (0.93%) for a between-group difference in the change in mortality that was less than the prespecified noninferiority margin of 1 percentage point. Differences in changes between the flexible programs and the standard programs in the unadjusted rate of readmission at 7 days, patient safety indicators, and Medicare payments were also below 1 percentage point; the noninferiority criterion was not met for 30-day readmissions or prolonged length of hospital stay. Risk-adjusted measures generally showed similar findings.

CONCLUSIONS—Allowing program directors flexibility in adjusting duty-hour schedules for trainees did not adversely affect 30-day mortality or several other measured outcomes of patient safety. (Funded by the National Heart, Lung, and Blood Institute and Accreditation Council for Graduate Medical Education; iCOMPARE ClinicalTrials.gov number, NCT02274818.)

FOR DECADES, THERE HAS BEEN DEBATE about the effects of the long duty hours of resident physicians, including questions concerning the safety of patients who are cared for by those trainees (both interns and residents), the education that trainees receive, and their sleep patterns and well-being. In an attempt to answer some of these questions, during the 2015–2016 academic year, we performed the iCOMPARE (Individualized Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education) trial in 63 internal-medicine residency programs in the United States. Residency programs underwent cluster randomization to a group with standard duty hours, as adopted by the Accreditation Council for Graduate Medical Education (ACGME) in July 2011, or to a group that permitted more flexible duty hours (principally, removing the 16-hour restriction on shift length).

The primary outcome of the iCOMPARE trial was patient safety, the results of which are reported here. The trial also assessed effects on trainee education,[1] with findings that showed no significant between-group difference in the proportion of time that interns spent on direct patient care or education but lower satisfaction with educational quality and overall well-being among those in the flexible programs. In addition, the trial assessed sleep patterns and alertness among interns, with findings that showed noninferiority of the duration of sleep in the flexible programs to that in the standard programs (see the article by Basner et al. in this issue of the *Journal*).[2]

With respect to patient safety, our primary hypothesis was that the unadjusted change in 30-day all-cause mortality[3–9] from the pretrial year to the trial year in the flexible programs would not be worse than that in the standard programs by more than 1 percentage point (noninferiority margin). We also evaluated five secondary non-inferiority hypotheses related to the outcomes of 7-day and 30-day rates of hospital readmission or death,[5,10] patient safety indicators (according to the Agency for Healthcare Research and Quality [AHRQ]),[11–14] prolonged length of hospital stay,[14–17] and payments made by Medicare. Each of these hypotheses had the same noninferiority margin of 1 percentage point.

METHODS

TRIAL OVERSIGHT

Details regarding the iCOMPARE trial have been reported previously.[1,18] The institutional review board at the University of Pennsylvania approved the protocol (available with the full text of this article at NEJM.org) and served as the institutional review board of record for all participating programs that signed on to an institutional affiliation agreement. Twenty-three programs opted for a local review process. The institutional review board at Children's Hospital of Philadelphia reviewed and approved the analysis of Medicare claims to test safety hypotheses.

TRIAL DESIGN AND PROGRAM SELECTION

A total of 63 programs underwent randomization in a 1:1 ratio to a group with standard duty-hour rules (following the 2011 ACGME duty-hour regulations with its 16-hour limit on intern shift length) or to a group with flexible duty hours, which allowed directors to extend work-hour limits beyond the 16-hour limit. (A summary of regulatory differences between the two duty-hour policy groups is provided in Table S1 in the Supplementary Appendix, available at NEJM.org.) In the flexible programs, directors selected services in which to implement the flexible rules and generally maintained flexible shifts in those services for the duration of the trial year.

We selected programs to meet sample-size requirements for the primary hypothesis. We included only programs with at least one affiliated hospital in both the upper half of resident-to-bed ratios and the upper three quartiles of patient volumes for 17 prespecified medical conditions (which included those chosen for their common treatment on internal-medicine services and their elevated mortality). Before randomization, at least one such hospital in each program had to be identified by the program director as being a hospital in which the director would implement flexible schedules if the program was randomized to the flexible-policy group. A total of 179 internal-medicine programs were approached, and 63 of their directors agreed to participate. The programs and the hospitals that were designated by program directors are listed in Table S2 in the Supplementary Appendix and constitute the trial populations. Complete Medicare data from all 63 programs were available for analysis.

PATIENT POPULATION AND DATA

All outcomes were ascertained from Medicare claims to ensure uniform measurement across participating hospitals.[14,19] Claims records were obtained from the Medicare Inpatient, Outpatient, Physician Part B, Home Health Agency, and Hospice files. We also used the Centers for Medicare and Medicaid Services Master Beneficiary Summary file for beneficiary demographic, vital status, and insurance information, and a validated date of death.[20] We selected claims for patients who were 65.5 years of age or older and who were admitted with one of the 17 qualifying medical conditions to a hospital in the iCOMPARE trial during the pretrial year (July 1, 2014, to June 30, 2015) or the trial year (July 1, 2015, to June 30, 2016). If a patient had multiple qualifying admissions, we included the first qualifying admission during each of the pretrial and trial years. Only patients who were in

Medicare fee-for-service for a period of at least 6 months before the index admission and at least 30 days after the index admission were included.

In October 2015, the *International Classification of Diseases, 10th Revision* (ICD-10) system was adopted by Medicare. All ICD-10 codes were recoded to those of the ninth revision of the ICD[21] (Table S3 in the Supplementary Appendix).

STATISTICAL ANALYSIS

In the primary analysis, we tested the hypothesis that the change in 30-day all-cause mortality from the pretrial year to the trial year in the flexible programs would not be worse than that in the standard programs by more than 1 percentage point (noninferiority margin). A secondary analysis examined risk-adjusted mortality. We also report the results of five secondary analyses of additional safety measures: rates of readmission or death within 7 days and 30 days after discharge, in which in-hospital deaths were counted as readmissions on discharge day zero (death date) to avoid inappropriate credit for an early death in the readmission analysis; the rate of at least one patient safety indicator, according to AHRQ criteria (Table S4 in the Supplementary Appendix); payments made by Medicare (Section S1 in the Supplementary Appendix); and the rate of a prolonged length of hospital stay.[14–17] A prolonged stay was defined as a condition-specific length of stay that exceeded the point in a hospitalization when rates of discharge typically begin to decline and was derived from lengths of stay at eligible hospitals not participating in the trial. For example, a prolonged stay for pneumonia was defined as a stay longer than 3 days (Table S5 in the Supplementary Appendix).

For each outcome, we report the average rate in each trial group in each year (trial year and pretrial year), the change from the trial year versus the pretrial year within each group, and the between-group difference in the change in the outcome from the trial year to the pretrial year (difference-in-difference analysis). The non-inferiority margin of 1 percentage point was chosen for each outcome, which would indicate a difference-in-difference not exceeding 1 percentage point and provide evidence that the flexible policy did not adversely affect patient outcomes as compared with the standard policy (Section S2 in the Supplementary Appendix). We used the t-test to compare groups with respect to continuous outcomes and the chi-square test for dichotomous outcomes. Our noninferiority trial did not include a prespecified plan for multiple comparisons but instead defined only one primary outcome (unadjusted 30-day all-cause mortality at any location) and five secondary outcomes, with risk-adjusted results reported as additional secondary outcomes.

The risk-adjustment models controlled for the qualifying medical conditions and coexisting conditions, as defined by Elixhauser et al.,[22,23] with some additional variables (Section S3 in the Supplementary Appendix). All covariates for patients were ascertained by means of a 6-month review of claims and present-on-admission logic as implemented in past studies.[7,14,17] The risk models also included age categories, sex, race or ethnic group, transfer-in status, hospice status, admission through emergency department, and direct admission to the intensive care unit (Table S6 in the Supplementary Appendix). All risk models were developed with the use of data from patients at 154 hospitals that met the criteria for inclusion in iCOMPARE but that were not affiliated with a randomized residency program

(Table S7 in the Supplementary Appendix). For risk adjustment, we performed regression modeling using PROC LOGISTIC[24] software for binary outcomes and PROC ROBUSTREG[25] software for continuous outcomes (SAS Institute). All analyses, unless specified, are based on an intention-to-treat approach.

Since flexible programs had the discretion to extend hours beyond the 2011 ACGME limits on any, all, or no services at their hospitals, we performed a subgroup analysis involving patients who were treated on services that were chosen to use flexible schedules and who were admitted for one of the 17 qualifying medical conditions. To identify these patients, we asked each program director to provide the dates when their attending physicians were supervising trainees on flexible services. We used the attending physicians' National Provider Identification Numbers and the Medicare Inpatient (Part A) and Physician Part B claims to identify the subgroup of trial-year patients who were on a flexible service on the first day of hospitalization for their index admission. In this analysis, patients who were treated by the same attending physician on the same services (intensive care unit or medical floor) in the same hospital during the pretrial year provided the pretrial data in the flexible programs. For the standard programs, we used the subgroup of patients in standard programs from either year who had an attending physician with data for the same services in both the pretrial and trial years. To ensure stable estimates for this analysis, we required each program to have at least 100 patients eligible for analysis in each of the pretrial and trial years. Since neither the attending physicians nor their patients underwent randomization to flexible programs within a hospital, we also report the average risk of death at the time of admission among patients who were included in this focused analysis and analyze the risk-adjusted outcomes (Tables S8 and S9 in the Supplementary Appendix).

RESULTS

PATIENTS

Of the 244,180 patients in the data set, 189,176 (77.5%) had one qualifying admission, 36,135 (14.8%) had two qualifying admissions, and 18,869 (7.7%) had three or more qualifying ad missions during the 2-year period. Using each patient's first admission per year, we studied a total of 264,585 admissions over the pretrial and trial years. The hospitals and patients in the two groups were very similar during the pretrial and trial years, with only a slight difference in the age of the patients, which suggests that the program randomization generally achieved balance in the types of hospital and patient characteristics (Table 1, and Table S10 in the Supplementary Appendix).

Table 2 describes the distribution of qualifying medical conditions among the patients included in the two groups. There were slight between-group differences in the distribution of some diagnoses, but a significant difference-in-difference between the two groups was observed only for renal failure.

OUTCOMES

The change in 30-day mortality (primary outcome) among the patients in the flexible programs (12.5% in the trial year vs. 12.6% in the pretrial year) was noninferior to that in

the standard programs (12.2% in the trial year vs. 12.7% in the pretrial year). The test for noninferiority was significant (P = 0.03), with an estimate of the upper limit of the one-sided 95% confidence interval (0.93%) for the between-group difference in the change in mortality that was less than the prespecified noninferiority margin of 1 percentage point (Table 3 and Fig. 1). (Unadjusted and risk-adjusted results for additional secondary outcomes are provided in Tables S11 and S12, respectively, in the Supplementary Appendix, with risk-adjusted results provided in Fig. S1.)

For the secondary outcomes, in the flexible programs, we observed noninferior results in risk-adjusted 30-day mortality, in both the un-adjusted and risk-adjusted analyses of 7-day re-admissions, AHRQ patient safety indicators, and Medicare payments; 30-day readmissions were noninferior in the risk-adjusted analysis. Of note, the low rate of patient safety indicators in the two groups makes a 1 percentage point noninferiority margin a generous standard, and the results should be interpreted with caution. The rate of a prolonged length of hospital stay did not meet the noninferiority margin, but with a baseline rate of 61% in the standard programs, a margin of 1 percentage point is highly conservative.

SUBGROUP ANALYSIS OF FLEXIBLE-SHIFT IMPLEMENTATION

The percentage of patients in flexible programs who could be definitively identified as having been admitted to a flexible service varied across programs. In the trial year, 61,194 patients were admitted to hospitals with flexible programs; of these patients, 51,813 were admitted to such hospitals that provided schedule information. (Of the 32 programs, 3 did not share schedule information for their hospitals.) We were able to link 15,977 patients in flexible programs (30.8%) to an attending physician who had ever supervised a flexible service, according to the schedule data provided by the program. Of those patients, 12,209 (76.4%) were admitted while the attending physician was supervising a flexible service, with the remaining 23.6% admitted when the attending physician was not supervising a flexible service.

After applying the criterion of the 100-patient minimum per year per program, we analyzed data from 10,459 patients who definitively had been exposed to a flexible schedule, as defined by a qualifying attending physician in a flexible program. (The number of patients and maximum shift lengths are provided in Table S8 in the Supplementary Appendix.) To account for observed differences in the severity of illness between groups (Table S9 in the Supplementary Appendix), we report only adjusted results in Figure 2. The adjusted analysis showed a difference in 30-day mortality of less than 1 percentage point, as did payments from Medicare. Additional outcomes for 7-day and 30-day readmissions showed some differences that slightly exceeded 1 percentage point, and the outcome of a prolonged length of hospital stay exceeded the non-inferiority margin, although, as noted, the noninferiority margin of 1 percentage point for a rate of more than 60% for a prolonged length of stay is very conservative. (Details regarding the unadjusted subgroup analysis are provided in Table S13 in the Supplementary Appendix for completeness, but since the implementation of a flexible schedule in a hospital service within a flexible program was not randomized, unadjusted results are suspect.) Although it is possible that some physician’s assistants participated in the care of these patients, we think that such staffing was rare; in each trial

group and on average, physician's assistants constituted less than 5% of the group of physician's assistants plus trainees at these hospitals.

DISCUSSION

The iCOMPARE trial was conducted to prospectively evaluate the implications of more flexible resident duty-hour rules on patient safety, trainee education, and intern sleep and alertness. The results of this trial, comparing outcomes in 32 flexible programs with 121,951 admissions with outcomes in 31 standard programs with 142,634 admissions during the pretrial and trial years combined, showed that there was no apparent harm to patients when programs were given more flexibility with duty-hour standards.

An accurate understanding of these findings depends critically on recognizing what our trial does and does not evaluate. It does not evaluate what happens when the trainees work extended shifts. Flexible programs in the trial were permitted but not required to use extended shifts. Although such programs used extended shifts for some rotations, all the programs also used standard shift lengths for others. That finding itself is revealing, in that the directors of the flexible programs did not use all their newly permitted latitude.

More than 30 years ago, duty-hour regulations came under intense scrutiny after the death of Libby Zion, an 18-year-old girl who died in a New York hospital in 1984 after being cared for by an intern and junior resident who had each been nearing the end of a long shift.[26] At that time, internal-medicine programs were relying on schedules with more extended cumulative and continuous duty hours. During the years of this trial, both policy groups of the iCOMPARE trial were bound by 80-hour weekly limits, minimum days off, and the strengthened 2011 ACGME supervision rules that aimed to make care safer in the era in which this trial was conducted.

Some observers may see the varied use of extended shifts as a weakness of the trial, one that dilutes the exposure of the intervention group. In contrast, we designed this feature as a strength of this pragmatic trial. Real policies set upper limits, not lower limits, on duty hours. The flexibility also recognizes the reality of participation of local stakeholders (including trainees) in schedule design. Indeed, we evaluated what happens today when program directors, perhaps better attuned to chronobiology and the interests of their trainees and patients than in previous eras, are permitted more flexibility than allowed by previous regulations. It turned out that patient safety was unchanged.

Another likely reason for this absence of differences in patient safety is that, for the ranges evaluated in this trial, shift lengths used in the programs are not extreme, with the result that effects on resident performance are minimal or mitigated by other safety processes. The subgroup analysis, which included only patients who were cared for by trainees who were serving on assigned flexible schedule services, also supported noninferiority. A companion article, which details the sleep and alertness of a subgroup of interns in this trial, shows that those who followed flexible schedules did not sleep for shorter periods than those in standard programs but did change their sleep patterns to compensate for longer shifts.[2]

Either way, this trial suggests that well-meaning regulations that were designed to correct past problems in hospitals need to be judged against their current relevance.

The trial design has several important strengths. By embedding a difference-in-difference analysis within a randomized trial, the design not only balanced both observed and unobserved patient and hospital characteristics across policy groups (which is characteristic of other randomized trials) but also ensured that the hospital environments in the trial year and pretrial year were similar for each randomized program. Results were confirmed with secondary risk-adjusted analyses to control for potential differences in patient populations in the two groups. The use of Medicare data allowed for uniform measurement of patient safety for all outcomes and allowed for the tracking of the rate of death from any cause at 30 days inside or outside the hospital. This factor is critically important, given that in-hospital mortality alone may be misleading when the length of hospital stay differs across hospitals because of different discharge policies and practices.

In conclusion, an analysis of patient outcomes from this cluster-randomized trial conducted in 63 internal-medicine residency programs across the United States suggests that allowing program directors the discretion to make their own schedules without continuous duty-hour limits did not result in worse patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by grants (U01HL125388, to Dr. Asch, and U01HL126088, to Dr. Tonascia) from the National Heart, Lung, and Blood Institute and grants from the Accreditation Council for Graduate Medical Education to Drs. Desai, Shea, and Silber.

We thank the participating program directors, along with Amanda K. Bertram, M.S., of Johns Hopkins University; Kelsey A. Gangemi, M.P.H., of the University of Pennsylvania; and Thomas Nasca, M.D., of the Accreditation Council for Graduate Medical Education.

REFERENCES

1. Desai SV, Asch DA, Bellini LM, et al. Education outcomes in a duty-hour flexibility trial in internal medicine. N Engl J Med 2018; 378: 1494–508. [PubMed: 29557719]
2. Basner M, Asch DA, Shea JA, et al. Sleep and alertness in a duty-hour flexibility trial in internal medicine. N Engl J Med 2019; 380: 915–23. [PubMed: 30855741]
3. Drye EE, Normand SL, Wang Y, et al. Comparison of hospital risk-standardized mortality rates calculated by using inhospital and 30-day models: an observational study with implications for hospital profiling. Ann Intern Med 2012; 156: 19–26. [PubMed: 22213491]
4. Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. JAMA 1997; 278: 2080–4. [PubMed: 9403422]
5. Patel MS, Volpp KG, Small DS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and re-admissions among hospitalized Medicare patients. JAMA 2014; 312: 2364–73. [PubMed: 25490327]
6. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. JAMA 2007; 298: 984–92. [PubMed: 17785643]

7. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. JAMA 2007; 298: 975–83. [PubMed: 17785642]
8. Volpp KG, Rosen AK, Rosenbaum PR, et al. Did duty hour reform lead to better outcomes among the highest risk patients? J Gen Intern Med 2009; 24: 1149–55. [PubMed: 19455368]
9. Volpp KG, Small DS, Romano PS, et al. Teaching hospital five-year mortalit trends in the wake of duty hour reforms. J Gen Intern Med 2013; 28: 1048–55. [PubMed: 23592241]
10. Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA 2013; 309: 587–93. [PubMed: 23403683]
11. Iezzoni LI, Daley J, Heeren T, et al. Using administrative data to screen hospitals for high complication rates. Inquiry 1994; 31: 40–55. [PubMed: 8168908]
12. McCarthy EP, Iezzoni LI, Davis RB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? Med Care 2000; 38: 868–76. [PubMed: 10929998]
13. Rosen AK, Loveland SA, Romano PS, et al. Effects of resident duty hour reform on surgical and procedural patient safety indicators among hospitalized Veterans Health Administration and Medicare patients. Med Care 2009; 47: 723–31. [PubMed: 19536029]
14. Silber JH, Rosenbaum PR, Kelz RR, et al. Medical and financial risks associated with surgery in the elderly obese. Ann Surg 2012; 256: 79–86. [PubMed: 22566017]
15. Silber JH, Rosenbaum PR, Even-Shoshan O, et al. Length of stay, conditional length of stay, and prolonged stay in pediatric asthma. Health Serv Res 2003; 38: 867–86. [PubMed: 12822916]
16. Silber JH, Rosenbaum PR, Koziol LF, Sutaria N, Marsh RR, Even-Shoshan O. Conditional length of stay. Health Serv Res 1999; 34: 349–63. [PubMed: 10199680]
17. Silber JH, Rosenbaum PR, Rosen AK, et al. Prolonged hospital stay and the resident duty hour rules of 2003. Med Care 2009; 47: 1191–200. [PubMed: 19786912]
18. Shea JA, Silber JH, Desai SV, et al. Development of the individualised Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education (iCOMPARE) trial: a protocol summary of a national cluster-randomised trial of resident duty hour policies in internal medicine. BMJ Open 2018; 8(9): e021711.
19. Navathe AS, Silber JH, Small DS, et al. Teaching hospital financial status and patient outcomes following ACGME duty hour reform. Health Serv Res 2013; 48: 476–98. [PubMed: 22862427]
20. Chronic Conditions Data Warehouse. Codebook: master beneficiary summary file — base with Medicare Part A/B/C/D. Version 1.1. January 2019 (https://www.ccwdata.org/documents/10280/19022436/codebook-mbsf-abcd.pdf).
21. Centers for Medicare & Medicaid Services. 2016 ICD-10-CM and GEMs. 2016 General Equivalence Mappings (GEMs) — diagnosis codes and guide. 2015 (https://www.cms.gov/Medicare/Coding/ICD10/2016-ICD-10-CM-and-GEMs.html).
22. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998; 36: 8–27. [PubMed: 9431328]
23. Healthcare Cost and Utilization Project. Elixhauser comorbidity software, version 3.7. Rockville, MD: Agency for Health-care Research and Quality, 2017 (https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp).
24. SAS Institute. Chapter 53: the LOGISTIC procedure In: SAS/STAT 9.3 user's guide. Cary, NC: SAS Institute, 2011: 4033–267 (https://support.sas.com/documentation/cdl/en/statug/63962/PDF/default/statug.pdf).
25. SAS Institute. Chapter 77: the ROBUSTREG procedure In: SAS/STAT 9.3 user's guide. Cary, NC: SAS Institute, 2011: 6531–625 (https://support.sas.com/documentation/cdl/en/statug/63962/PDF/default/statug.pdf).
26. Asch DA, Parker RM. The Libby Zion case. N Engl J Med 1988; 318: 771–5. [PubMed: 3347226]

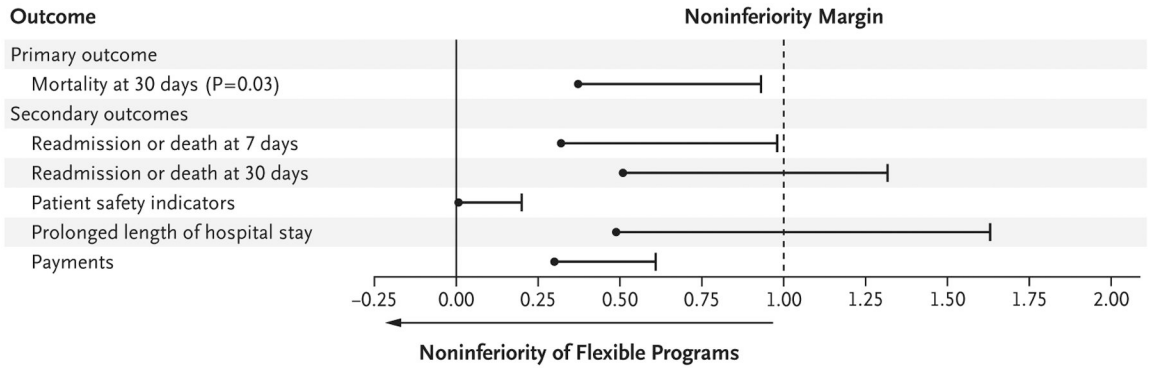

Figure 1. Patient Safety Outcomes.

Shown are one-sided 95% confidence intervals for the difference between flexible programs and standard programs in the primary and secondary outcomes for patient safety. An outcome in the flexible programs was deemed to be noninferior to that in the standard programs if the upper limit of the confidence interval for the difference (the value in the flexible programs minus the value in the standard programs) was less than the noninferiority margin of 1 percentage point. Confidence intervals for binary outcomes represent the absolute difference in the outcome between the trial year minus the pretrial year in the flexible programs minus the corresponding absolute difference in the standard programs. Confidence intervals for payments represent the percent change (trial year vs. pretrial year) in the flexible programs minus the percent change in the standard programs with the use of log transformation. Patient safety indicators were determined according to the criteria of the Agency for Healthcare Research and Quality. Details regarding the cutoff points for determining whether a hospital stay has a prolonged length for various medical conditions are provided in Table S5 in the Supplementary Appendix.

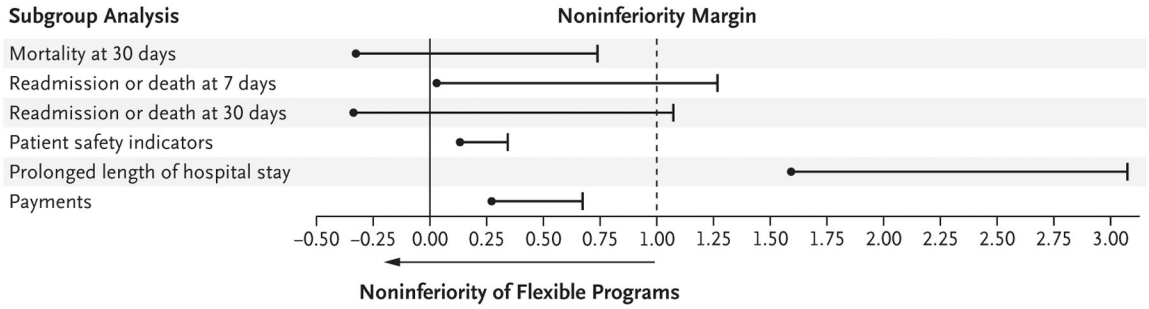

Figure 2. Subgroup Analysis of Flexible-Shift Implementation.

Shown are the results of a subgroup analysis of data from the patients in Figure 1 who were treated on services that were specifically designated for flexible-shift implementation by the program director during the trial year; all the pretrial-year patients in this subgroup had been treated by the same attending physicians who worked on the flexible services during the trial year. In the standard programs, all the patients of attending physicians who were working during both years were included. To maximize the stability of the estimates, only programs with a minimum of 100 eligible patients in each year (20 flexible programs and 30 standard programs) were included in the analysis. All analyses are risk-adjusted for coexisting medical conditions, as described in Table S6 in the Supplementary Appendix. An outcome in the flexible programs was deemed to be noninferior to that in the standard programs if the upper limit of the confidence interval for the difference (the value in the flexible programs minus the value in the standard programs) was less than the noninferiority margin of 1 percentage point.

Table 1.

Characteristics of Hospitals and Patients in the Flexible and Standard Programs.*

Characteristic	Flexible Programs (N = 32)		Standard Programs (N = 31)		Between-Group Difference†
	Trial Year	Pretrial Year	Trial Year	Pretrial Year	
Hospitals‡					
Mean resident-to-bed ratio	0.66	NA	0.57	NA	0.09
Mean no. of beds	600.9	NA	560.3	NA	40.6
Mean ratio of nurses to beds	1.71	NA	1.47	NA	0.24
Ratio of RNs to RNs plus LPNs	0.95	NA	0.96	NA	−0.01
	(N = 61,194)	**(N = 60,757)**	**(N = 71,662)**	**(N = 70,972)**	**Difference in Change**
Patients					
Mean age (yr)	77.7	77.9	78.6	79.0	0.2§
Male sex (%)	50.7	50.0	49.1	48.6	0.2
Race or ethnic group (%)¶					
Non-Hispanic white	74.5	74.8	78.6	79.0	0.1
Black	20.2	20.3	15.7	15.8	<0.1
Hispanic	1.8	1.7	1.4	1.3	−0.1
Other	3.4	3.2	4.2	3.9	<0.0
Admission status (%)					
Emergency	13.6	13.2	11.7	11.2	−0.1
Transfer	7.3	7.1	5.5	5.4	<0.1
Risk of death at 30 days	12.7	12.6	12.4	12.4	0.1
Medical conditions (%)					
Congestive heart failure	42.3	43.1	41.6	42.5	0.1
Diabetes					
Uncomplicated	42.2	43.7	40.5	41.5	−0.5
Complicated	28.0	21.4‖	26.5	20.7‖	0.8
Chronic pulmonary disease	39.5	39.5	38.8	38.7	−0.2
Renal failure	37.9	37.2	36.7	36.2	0.1
Peripheral vascular disease	32.9	32.7	34.3	33.6	−0.5
Valvular heart disease	27.3	26.9	27.8	27.8	0.4

[*] Shown are the average numbers in each group in the pretrial year (July 1, 2014, to June 30, 2015) and the trial year (July 1, 2015, to June 30, 2016), along with the between-group difference in the change in the characteristic between the pretrial year and the trial year (difference-in-difference analysis). Details regarding additional characteristics of the hospitals and patients are provided in Table S10 in the Supplementary Appendix. LPN denotes licensed practical nurse, NA not applicable, and RN registered nurse.

[†] Values that are shown for differences and difference-in-differences are for the flexible programs, as compared with the standard programs.

[‡] At the beginning of the trial year, hospital characteristics were evaluated in 35 facilities in the flexible programs and in 38 facilities in the standard programs because some programs had multiple affiliated hospitals (Table S2 in the Supplementary Appendix).

[§] $P<0.01$.

[¶] Race or ethnic group was reported to Medicare by the patients.

[‖] $P<0.001$.

Table 2.

Medical Conditions Qualifying Patients for Inclusion in Trial.*

Principal Diagnosis Group	Flexible Programs		Standard Programs		Difference in Change
	Trial Year (N-61,194)	Pretrial Year (N = 60,757)	Trial Year (N = 71,662)	Pretrial Year (N = 70,972)	
	percent				*percentage points*
Septicemia	18.3	17.2	18.0	16.7	−0.2
Congestive heart failure	13.0	12.7	12.5	12.4	0.2
Stroke	12.3	11.6	11.9	11.3	0.2
Acute myocardial infarction	9.3	9.5	7.9	7.5	−0.5
Coronary atherosclerosis	6.5	6.2	5.4	5.4	0.2
Gastrointestinal bleeding	5.7	5.7	6.3	6.4	0.1
Cardiac arrhythmia	5.7	6.5	5.1	5.6	−0.3
Renal failure	5.3	5.7	5.6	5.6	−0.4†
Pneumonia	4.7	5.0	5.8	6.6‡	0.5
Chronic obstructive pulmonary disease, asthma, or bronchitis	3.9	4.5	5.2	5.5	−0.3
Pulmonary embolism	3.7	3.7	3.6	3.4	−0.2
Acute respiratory disorder	3.3	3.3	2.6	2.7	0.2
Cellulitis	2.6	2.6	3.4	3.4	<0.1
Syncope	1.9	2.0	2.3	2.5	0.1
Chest pain	1.5	1.6	1.6	1.8	0.1
Intestinal infection	1.2	1.3	1.5	1.8	0.2
Acute pancreatitis	1.1	1.0	1.2	1.2	0.0

*Details regarding principal diagnosis codes in the *International Classification of Diseases, 9th Revision* (ICD-9) and ICD-10 are provided in Table S3 in the Supplementary Appendix. The t-test was used to assess the difference between the pretrial year and the trial year in the mean percentage of patients who had each qualifying condition within each group and to assess the between-group difference in the change between trial years.

†P<0.01.

‡P<0.05.

N Engl J Med. Author manuscript; available in PMC 2019 September 07.

Table 3.

Patient Safety Outcomes.*

Outcome	Flexible Programs (N = 32)	Standard Programs (N = 31)	Difference in Change (95% CI)†
Primary outcome			
30-day mortality			
Trial yr (%)	12.5	12.2	0.3
Pretrial yr (%)	12.6	12.7	−0.1
Difference in percentage points	−0.1	−0.5	0.4 (−∞ to 0.9)
Secondary outcomes			
Readmission or death at 7 days			
Trial yr (%)	16.9	16.6	0.3
Pretrial yr (%)	16.6	16.7	0.0
Difference in percentage points	0.3	−0.1	0.3 (−∞ to 1.0)
Readmission or death at 30 days			
Trial yr (%)	29.9	29.3	0.7
Pretrial yr (%)	29.8	29.7	0.1
Difference in percentage points	0.1	−0.4	0.5 (−∞ to 1.3)
Patient safety indicators‡			
Trial yr (%)	0.9	0.7	0.2
Pretrial yr (%)	1.0	0.7	0.2
Difference in percentage points	−0.1	−0.1	<0.1 (−∞ to 0.2)
Prolonged length of hospital stay§			
Trial yr (%)	63.2	61.2	2.0
Pretrial yr (%)	63.0	61.4	1.5
Difference in percentage points	0.3	−0.2	0.5 (−∞ to 1.6)
Payment in 2016 dollars¶			
Trial yr	25,139	23,199	1940
Pretrial yr	23,882	21,870	2012
Relative difference (%)	0.7	0.5	0.3 (−∞ to 0.6)

*All listed values are means.

[†] One-sided 95% confidence intervals (CIs) were calculated to complement tests of noninferiority. If the upper limit of the confidence interval for the value in the flexible programs minus that in the standard programs was less than the noninferiority margin of 1 percentage point, an outcome in the flexible programs was deemed to be noninferior to that in the standard programs. Confidence intervals have not been adjusted for multiple testing, so inferences drawn from the intervals may not be reproducible.

[‡] Patient safety indicators include rates of pressure ulcers, iatrogenic pneumothorax, bloodstream infection from a central venous catheter, hip fracture, hemorrhage or hematoma, physiologic or metabolic derangement, respiratory failure, pulmonary embolism or deep-vein thrombosis, sepsis, and accidental puncture or laceration. Details are provided in Table S4 in the Supplementary Appendix.

[§] A prolonged length of hospital stay is defined as a length of stay that exceeded the point at which the rate of discharge typically begins to decrease. Details regarding the number of days defining prolonged length of stay for each condition are provided in Table S5 in the Supplementary Appendix.

[¶] For clarity, the mean dollars in the trial year and pretrial year are listed without the use of log transformation. Because payment data are skewed, log-transformed dollars were used to calculate the relative percent differences for each program, which were then aggregated to each trial group. The associated 95% confidence interval was again based on log-transformed dollars. Formulas for this calculation are provided in Sections S1 and S2 in the Supplementary Appendix.